

De novo сборка транскриптома лиственницы сибирской (*Larix sibirica* Ledeb.)

Бирюков Владислав Владимирович

Студент (бакалавр)

Сибирский федеральный университет, Красноярск, Россия

E-mail: vladbir2010@gmail.com

Последние достижения в области технологий секвенирования нового поколения и разработки биоинформатических инструментов привели к возможности расшифровки нуклеотидных последовательностей больших геномов, а также стали реальными секвенирование и анализ транскриптома немодельных организмов, например, хвойных растений, которые в отличие от модельных организмов имеют такие признаки, как медленный рост, удивительно высокая модификационная изменчивость и огромный размер генома, который богат повторяющимися последовательностями, в основном, в виде транспозонов и tandemных повторов. Из-за этих особенностей процесс сборки нуклеотидной последовательности очень усложняется.

В данной работе производилась сборка данных секвенирования мРНК почек лиственницы сибирской (*Larix sibirica* Ledeb.), полученных в Лаборатории лесной геномики СФУ под руководством проф. К. В. Крутовского на приборе Illumina MiSeq. Была проведена предварительная обработка данных секвенирования с использованием таких инструментов, как Trimmomatic и FastQC. Сборка осуществлялась при помощи программного обеспечения Trinity и Velvet-Oases, качество полученной сборки оценивалось при помощи PRINSEQ и QUAST. Bowtie2 и SAMtools послужили вспомогательными инструментами.

В результате работы, была получена первичная сборка транскриптома лиственницы сибирской. Длина собранной нуклеотидной последовательности составила 26493048 п.н.о.. Для оценки ее качества, было произведено картирование ридов обратно на сборку (общий процент выравнивания составил 87.39%, среднее и максимальное покрытие - 23 и 8014 соответственно), а также были определены основные метрики: общее число контигов - 43686, из которых 6919 - контиги, длиной более 1000 п.н.о., длина максимального контига составила 8512 п.н.о. и N50 - 878 п.н.о. Сравнительный анализ нашей сборки с транскриптомами других хвойных, изученных ранее в работах других авторов, показал достаточно высокую степень сходства полученных контигов и контигов предыдущих сборок. Таким образом, на данный момент получена сборка, составляющая около 70% от предполагаемой общей длины транскриптома лиственницы сибирской.

Актуальность работы заключается в том, что сборка транскриптома играет важную роль в изучении регуляции экспрессии генов и регуляторных механизмов клетки и организма в целом, что является одной из актуальных проблем современной биологии.

Работа выполнена в рамках проекта «Геномные исследования основных бореальных лесобразующих хвойных видов и их наиболее опасных патогенов в Российской Федерации», финансируемого Правительством РФ (договор №14.Y26.31.0004).